Response to Referees on egusphere-2023-1556

We appreciate the reviews and comments from both Referees. Please find the response to Referee 2 on pages 1-8, and the response to Referee 3 on pages 9-10.

Response to Referee 1 on egusphere-2023-1556

First, we would like to thank the Referee for reviewing and commenting on the manuscript, which will improve the quality of the manuscript. Please find the item-by-item reply below, with the original comments in *italics* and the responses in blue. All the suggested changes are implemented in the revised manuscript.

The authors provided detailed replies to my comments, thank you. The manuscript has been significantly modified, and most contents improved. However, some of the issues pointed out in the previous round still hold after the modifications.

The main concern is still related to the investigated depth. Although it has been increased from the former 4 cm, 12 cm is still a too tiny layer especially at C- band - see e.g. Surdyk 2002 (10.1016/S0034-4257(01)00308-X), Picard et al. 2009 (10.3189/002214309788816678), Champollion et al. 2019 (10.5194/tc-13-1215-2019), Brucker et al., 2009 (10.3189/002214310792447806). I'm still convinced you should provide a physical explanation supporting this choice.

We understand the concern of the reviewer, and agree that indeed the penetration depths can exceed 1 m for 19 GHz and for C-band, as mentioned by e.g. Surdyk (2002) and Fraser et al. (2016). Meanwhile, such ranges of penetration depths are not always certain, e.g. in Picard et al. (2009), *"the deepest penetrations at the 19V channel are located in Marie Byrd Land (4–7 m) and on the East Antarctic divide (4–6 m). The shallowest penetrations are found in the wind-glazed surface regions and megadunes, with values as low as 0.3 m. Intermediate values are found in Wilkes Land between the coast and the divide (2.5–5 m)."* Furthermore, according to Arndt and Haas (2019), although the penetration depth of C-band exceeds 1 m, *"however, increased backscatter along the propagation path through the snow at any depth will result in the observed overall backscatter increases,"* and according to Cartwright et al. (2022), *"azimuthal anisotropy arises primarily due to the interaction between the incident microwave radiation and regularly aligned roughness (on the Rayleigh roughness scale, or larger) of the surface and subsurfaces within the penetration depth (Ulaby et al., 1996; Bingham and Drinkwater, 2000; Partington and Flach, 2003; Yurchak, 2009; Fraser et al., 2016),"* therefore, the shallower depth firn properties should not be completely negligible to long-wavelength microwave.

With the aforementioned reference, we understand it is true that ASCAT in principle is sensitive to the scattering properties up to several metres' depths, as that is indeed the estimated penetration of the C-band wavelength. However, there is no evidence that the relationship between the density at those depths and the C-band radar backscatter is a strong one; although a few modelling efforts have been carried out in that direction, there is no literature showing that the physical modelling is mature for active microwave sensing. Therefore, the contribution of C-band radars to the retrieval of snow properties at a wide range of depths remains an interesting aspect to investigate, especially through machine learning techniques which are more data-driven, including assimilation with other sensors.

We applied the approach conducted at 12 cm also to the snow density at other depths, and added the figure to this document (Fig. R1). The figure shows that both the RMSE and correlation coefficient reduce with an increasing depth. We believe that this is due to the fact that the firn density at larger depths is not largely influenced by surface temperature and precipitation, which have a larger impact on temporal variations of microwave signals.

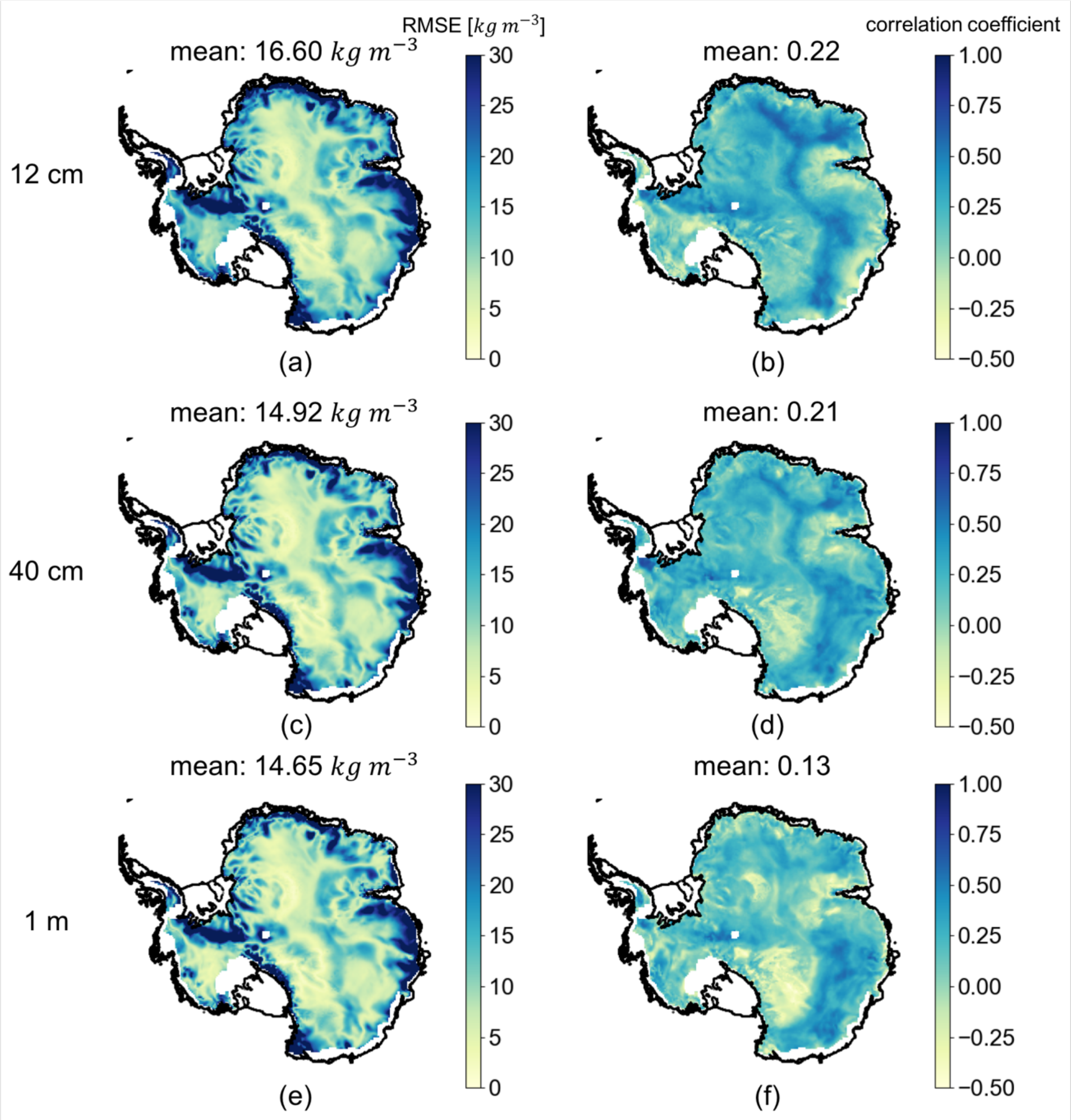

Figure 1. RMSE (left) and correlation coefficients (right) at different depths (12 cm, 40 cm and 1 m, respectively).

Furthermore, we also understand the concern that the microwave signals can be affected by layers that are at a larger depth (i.e. even larger than 1 m). Therefore, we computed the mean temporal correlation coefficients between densities at different depths and satellite parameters, summarised in Table R1. The table shows that within our available dataset, the correlation between all satellite parameters and densities on average reaches the maximum at 40 cm depth. It is then worth noting that the correlation first drastically decreases (by 40 %) between 40 cm and 1 m depths for 37 GHz, and then largely decreases (by 21 %) between 1 m and 2 m for 19 GHz. Therefore, we have adopted the density from 40 cm depth in the revised manuscript, and added Table R1 to the revised manuscript. It is important to note that since all depths within the penetration depth affect the scattering properties, the “density at depth…” should be the mean density of the upper x depths (x = 12 cm, 40 cm, 1m, etc.).

Table R1. Correlation coefficients between IMAU-FDM densities at different depths and satellite parameters.

	TB(19V)	TB(19H)	TB(37V)	TB(37H)	sigma^0
12 cm	0.19	0.18	0.20	0.20	-0.05
40 cm	0.24	0.23	0.20	0.19	-0.06
1 m	0.23	0.20	0.12	0.12	-0.06
2 m	0.18	0.12	0.03	0.02	-0.06
5 m	0.08	0.02	-0.07	-0.08	-0.04
10 m	0.05	0.01	-0.07	-0.07	-0.03

Also, the concern about redundant information in coupling microwave observations with their combinations in the RF retrieval, according to information theory, has not been solved: my point is confirmed by the predictor importance analysis in fig.6 that clearly shows the dominant role of direct observations and the minor role of derived indices.

The original purpose of using the brightness temperature ratios was to reproduce the Champollion et al. (2013) study, where the ratios could be related to near-surface hoar-crystal formation. But it is also true that we could not reproduce the method in our study, especially because it was not applicable to the entire Antarctic ice sheet. Hence, we agree to remove the derived parameters. The revised manuscript is now based on the new setting.

Another concern is about the SSM/I derived indices FR and PR: beside the reason for using the formulation in eq. 1 and 2 instead of the other ratio generally adopted (e.g. Kelly et al., 2003; Tedesco et al., 2004, Chang et al., 1987; Chang et al., 1990; Santi et al. 2012), the PR correlation with the target parameter does not seem exceptional (especially at Ku band) and the reasons because it is sometimes positive and sometimes negative quite is difficult to explain because it does not seem related to environmental factors as for the other observables shown in the figure.

We have removed the ratios in the revised manuscript.

I believe further revision should be done to clarify the compensation for observation angle: σ° is universally adopted to refer to backscattering, which is derived from the NRCS that already accounts for incident angle. Please also mention backscattering when introducing σ° notation.

It is true that σ^0 already includes the normalisation by the area of the cell resolution. However, this extra normalisation done in the BYU product also accounts for the differences in the physical response of the distributed target per metre square which is a function of the incidence angle. This is also documented in the methodology paper of Long and Drinkwater (2000):”Because scatterometers make measurements over a range of incidence angles, the incidence angle dependence of σ^0 must be accounted for.” To avoid confusion while being coherent with the A parameter as defined in Long and Drinkwater (2000) and used by Fraser et al. (2016), we have changed the σ^0 into σ_A^0 in the revised manuscript.

Total data, training and test datasets still need to be clearly quantified: I did not find the numbers as declared in the authors replies (possibly my fault?). At lines 264-269 it is stated *that Subset I contains 10% of the non-melting pixels (how many in total?) and Subset II is composed of 100 pixels (only?). Nothing is declared about independence of the two datasets. These numbers seem significantly smaller than my deductions for the previous round and this could lead to some over-dimensioning of RF parametrization shown in Table 1 (and consequent overfitting) Moreover they do not seem consistent with the amount of data arguable from fig. 6 left and even insufficient for generating the maps in figure 3 and 8. I suspect some misunderstanding.*

Regarding this problem, we agree that the numbers declared in the previous replies have not been added to the manuscript, and now we have added the numbers in the revised version. We would like to clarify that each pixel we use consists of time series of all parameters, i.e. 366 density estimations. Therefore, Subset I used for hyperparameter tuning and training consists of 1748*366 samples (instead of 1748 points), and Subset II used for testing and importance computation, i.e. Fig. 6 consists of 100*366 samples (instead of 100 points). We have double-checked that Subset I and Subset II do not have overlapping pixels. Regarding the amount of data in Fig. 6, we presented the testing result using Subset II, which consists of 100*366 samples. This is because the entire dry-snow zones in Antarctica according to our clustering method consists of 17478 pixels, and subsequently 17478*366 samples, which is too large and redundant for visualising in a scatter plot such as Fig. 6. However, for Fig. 7 and Fig. 8, we used the densities of the 17478*366 samples, and calculated the resulting mean densities and errors. The amount of data has been added to the revised manuscript lines 253—258.

Table I. In my personal experience, increasing the number of trees above 50 does heavily affect the computational cost without providing accuracy improvements. But this is just my experience, not absolute truth. In any case the most important thing to assess is the RF dimensioning in terms of training data amount (see comment above).

Please refer to the comment above. We believe that by using 1764*366 samples as training data, the over dimensioning problem should be resolved.

Figure 1 is useful addition, in my view however, it is difficult to understand in the current implementation: I would suggest revising and simplify.

This has been changed in the revised manuscript (also with the ratios removed).

Evaluating quantitatively the results in figure2 with respect to the corresponding figure in the former manuscript is not straightforward, however it seems that, by comparing against 12 cm rather than 4 cm, some small improvements are obtained at Ku and Ka band while the C- band does not show appreciable improvements. This could go in the right direction by supporting the concern about the insufficient depth. As requested in the previous round, it would be important to provide overall correlation or determination coefficients (at least for each cluster - choice is up to you, but you should be consistent through the manuscript) also to understand which is the contribution of RF with respect to the direct correlation between single observables and target parameters. The physical reasons supporting changes from positive to negative correlations should be better discussed in any case.

The correlation coefficients have been added to the revised manuscript. We have also attached a comparison between RF and a simple linear regression (Fig. 7 and lines 353—355).

Figure 5 seems a bit redundant and its informative content not exceptional since the behaviours are difficult to interpret; moreover, figure 6 points out the minor contribution of these parameters in the retrieval.

Figure 5 was originally provided with the purpose of demonstrating how complicated the relationship between satellite parameters and densities can be for dry snow, and that our clustering method could distinguish melt moments and the spatial coverage of melts. However, we agree that the informative content is not exceptional, hence will move it to appendix in case some readers may be curious about how the distinction of melt regions looks like. We also agree that these parameters are not contributive to the RF approach, hence they will be removed from the RF sections.

Figure 6 left. The scatterplot seems slightly improved wrt the former result at 4 cm, however I still see some saturation in the retrieved density: in my experience this could depend on not proper dimensioning/training of RF, any explanation? Again, how many data in the scatterplot? The requested correlation coefficient is not provided in the figure/caption.

The number of training data is 1748*366=639,768, and the number of testing data visualised in the scatterplot is 100*366 = 36,600. This information has been added to the revised manuscript.

Regarding the saturation in the retrieved density, we noticed that not only does RF largely underestimate the density higher than 410 $kg\ m^3$, it also overestimates the density lower than 325 $kg\ m^3$. Therefore, we presume that both the highest and the lowest IMAU-FDM densities are not properly accounted for, both due to uncertainties from the IMAU-FDM modelling process and the limitation of the combination of satellite parameters. An example is shown in Fig. R2 of this document. The scatterplot in the upper panel is calculated as the temporally averaged IMAU-FDM density subtracted by the temporally averaged RF density at the 100 sample pixels. On average, the largest underestimation for RF (exceeding 20 $kg\ m^3$) occurs in regions where the summer wind velocity is more than 2.5 $m\ s^{-1}$ lower than the winter wind velocity. This corresponds to part of our conclusion that IMAU-FDM shows pronounced different behaviours from the satellite time series when the seasonal wind velocity difference is high, hence IMAU-FDM may not adequately capture the actual physical meteorological phenomena that affect microwave scattering properties. On the other hand, RF on average largely overestimates the density in Transantarctic Mountains, potentially due to the complex terrain that affect the surface scattering of the microwave (instead of volume scattering). Finally, an overall limitation of using purely satellite data time series is that they are largely dependent on surface and near-surface temperature. Our study is therefore coherent with the Fraser et al. (2016) study, who could establish a relationship between long-term mean ASCAT backscatter and snow or climate properties for Antarctic dry snow, although our work focuses on establishing a relationship the other way around, i.e. re-constructing firn density using a combination of satellite data. However, the seasonal correlation is compromised both in the Fraser et al. (2016) study and in our study.

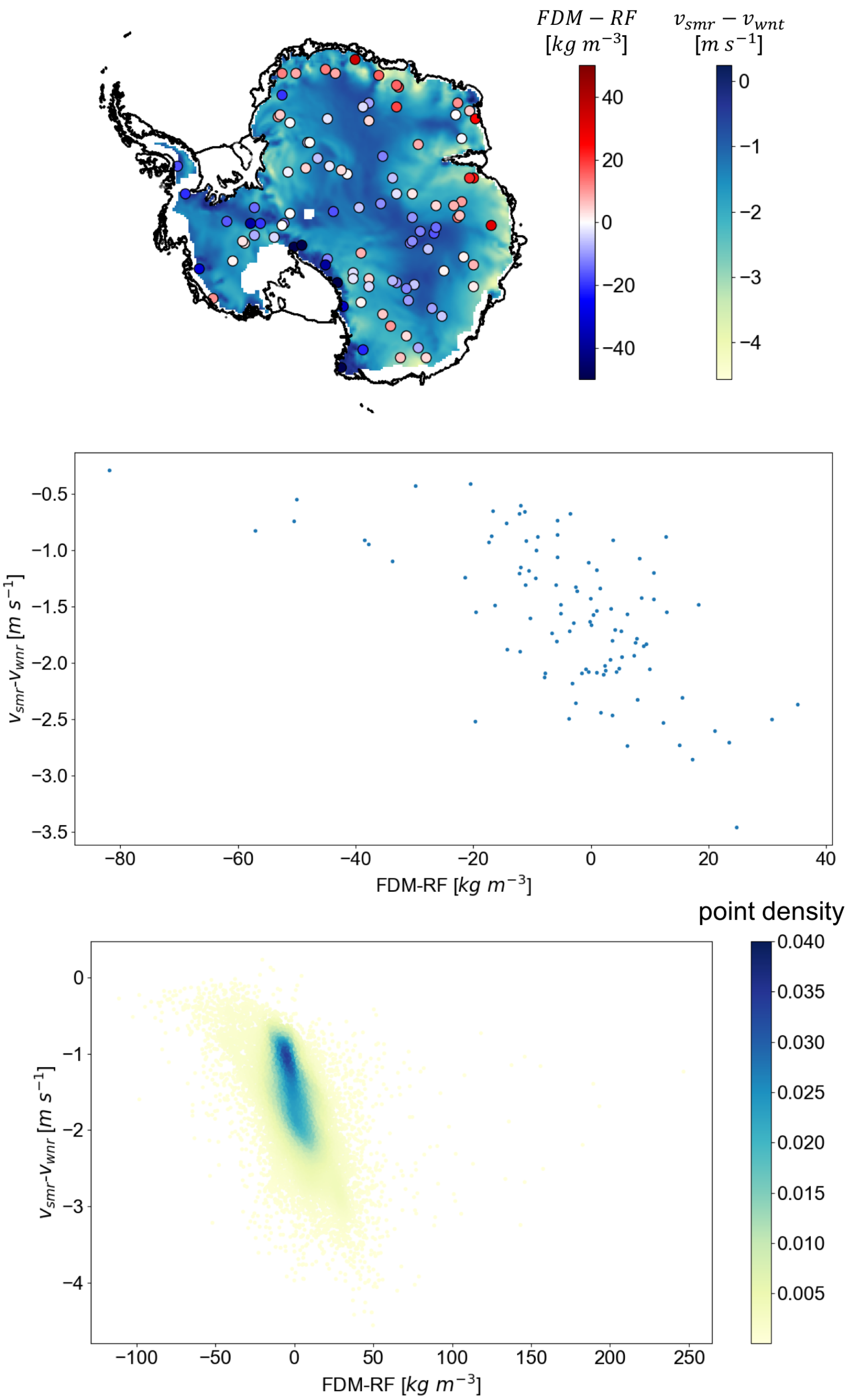

Figure R2. Temporal mean difference between the IMAU-FDM density and RF density at the 100 sample pixels, overlayed on the seasonal wind velocity difference map (upper); scatterplot of the temporal mean difference between the IMAU-FDM density and RF density at the 100 sample pixels versus the seasonal wind velocity at the sample pixels (middle); and temporal mean difference between the IMAU-FDM density and RF density at all pixels in dry snow zones versus seasonal wind velocity difference, coloured by the density distribution of points (lower).

Figure 6 right. thanks for the explanation about the Breiman vs. Gini predictor importance analysis, however showing one or another histogram is enough, also because they bring some contradictory results that are difficult to justify, up to you….

We have removed the ratios and anomalies and kept only the Gini predictor.

The comparison with the Dome-C data from Leduc is relevant as validation against independent data. However, the data refer to the first 2/3 cm depth and the RF has been trained for 12 cm depth…. Please further address.

This comparison was performed between the first 2 cm depth of field measurements and the 4 cm depth of IMAU-FDM. Then, both 4 cm and 12 cm IMAU-FDM densities are provided to show the large discrepancies between the model and the field measurement. Furthermore, the Dome C data from Leduc-Leballeur et al. was used also to analyse the potential correlation between near-surface hoar-crystal formation and disappearance and polarisation ratios. This is not applicable in our study anymore hence has been removed.

Figure 8 and discussion. What I would see addressed (see my comment in the previous round), is the comparison between the R coefficients shown here and those of single observables in figure 3, with the aim of pointing out the improvement brought by RF with respect of attempting the direct retrieval from single observables, that in some cases already reach very high correlation.

This has been added to the revised manuscript (Fig 7). However, we also show that RF outperforms the simple linear regression in terms of RMSE, which is an important indicator apart from the correlation coefficient.

Figure 9 and discussion. The rationale of showing the PR at Dome C after showing the overall RF performances as R and RMSE maps of the entire Antarctica is unclear to me. PR is just one of the inputs of the RF algorithm and not even one of the most important.

Originally, we referred to the Champollion et al. study where they attributed the variation in polarisation ratios to the hoar-crystal disappearance, which is characterised by an increase in near-surface density and a reduction in grain size. We understand this inclusion of the Champollion et al. study causes confusion, hence removed this part in the revised manuscript.

Reference

Arndt, S. and Haas, C.: Spatiotemporal variability and decadal trends of snowmelt processes on Antarctic sea ice observed by satellite scatterometers, The Cryosphere, 13, 1943–1958, https://doi.org/10.5194/tc-13-1943-2019, 2019.

Bingham, A. W. and Drinkwater, M. R.: Recent changes in the microwave scattering properties of the Antarctic ice sheet, IEEE T. Geosci. Remote, 38, 1810–1820, https://doi.org/10.1109/36.851765, 2000.

Cartwright, J., Fraser, A. D., and Porter-Smith, R.: Polar maps of C-band backscatter parameters from the Advanced Scatterometer, Earth Syst. Sci. Data, 14, 479–490, https://doi.org/10.5194/essd-14-479-2022, 2022.

Champollion, N., Picard, G., Arnaud, L., Lefebvre, E., and Fily, M.: Hoar crystal development and disappearance at Dome C, Antarctica: observation by near-infrared photography and passive microwave satellite, The Cryosphere, 7, 1247–1262, https://doi.org/10.5194/tc-7-1247-2013, 2013.

Fraser, A. D., Nigro, M. A., Ligtenberg, S. R. M., Legrésy, B., Inoue, M., Cassano, J. J., Kuipers Munneke, P., Lenaerts, J. T. M., Young, N. W., Treverrow, A., van den Broeke, M., and Enomoto, H.: Drivers of ASCAT C band backscatter variability in the dry snow zone of Antarctica, Journal of Glaciology, 62, 170–184, https://doi.org/10.1017/jog.2016.29, 2016.

Long, D. and Drinkwater, M.: Azimuth variation in microwave scatterometer and radiometer data over Antarctica, IEEE Transactions on Geoscience and Remote Sensing, 38, 1857–1870, https://doi.org/10.1109/36.851769, 2000.

Partington, K. and Flach, D.: Synergetic Use of Remote Sensing Data in Ice Sheet Snow Accumulation and Topographic Change Estimates: Comparison of model output with available data, Tech. Rep. NOV-3137-NT-1537, Noveltis, Vexcel UK and Legos, Ramonville-Saint-Agne, France, 2003.

Picard, G., Brucker, L., Fily, M., Gallée, H., and Krinner, G.: Modeling time series of microwave brightness temperature in Antarctica, Journal of Glaciology, 55, 537–551, https://doi.org/10.3189/002214309788816678, 2009.

Surdyk, S.: Using microwave brightness temperature to detect short-term surface air temperature changes in Antarctica: An analytical approach, Remote Sensing of Environment, 80, 256–271, https://doi.org/10.1016/s0034-4257(01)00308-x, 2002.

Ulaby, F. T., Siquera, P., Nashashibi, A., and Sarabandi, K.: Semi-empirical model for radar backscatter from snow at 35 and 95 GHz, IEEE T. Geosci. Remote, 34, 1059–1065, https://doi.org/10.1109/36.536521, 1996.

Yurchak, B.: Some Features of the Volume Component of Radar Backscatter from Thick and Dry Snow Cover, in: Advances in Geoscience and Remote Sensing, edited by: Jedlovec, G., Intech, Rijeka, Croatia, https://doi.org/10.5772/8339, 2009.

Response to Referee 3 on egusphere-2023-1556

First, we would like to thank the Referee for reviewing and commenting on the manuscript, which will improve the quality of the manuscript. Please find the item-by-item reply below, with the original comments in *italics* and the responses in blue. All the suggested changes are implemented in the revised manuscript.

This paper used machine learning (ML) and satellite microwave data to examine Antarctic firn density. The authors did a good job of responding to all comments in the first round of review. I only have minor comments.

Across the document, the notation for figure should be normalized (Figure, fig, or fig.). Should be Fig. everywhere.
We have corrected the manuscript for consistency. However, according to the guidelines of The Cryosphere (https://www.the-cryosphere.net/submission.html, Figures & tables section), we did not change everything:
"The abbreviation 'Fig.' should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence, e.g.: 'The results are depicted in Fig. 5. Figure 9 reveals that...'."

Line 74: Remove on. "Antarctic ice sheet based on on daily"
This has been corrected in the revised manuscript.

Line 155: What about the penetration depth of C-band? The Sigma0 at C-band is sensitive to deeper than this.
We conducted an experiment computing the temporal correlation between satellite observations and IMAU-FDM densities (Table 2 of the revised manuscript), and determined that the optimal depth where the density can be correlated with all frequencies should be 40 cm. We also refer to the Fraser et al. (2016) study where the top 1 m density was seen as the layer most sensitive to atmospheric drivers which can be correlated with C-band backscatter, despite a much larger penetration depth of C-band radar.

Line 174: replace "parameterization" to "parametrization"
This has been corrected (and other identical errors) in the revised manuscript.

Line 208: Is it spatial or time clustering? Or both? You talked about spatial and then...
It is a clustering of time series. In doing this, the regions which experienced intensive melt and subsequent ice-layer formation can be distinguished. To avoid confusion, the previous "spatial" has been removed.

Line 294: I suggest talking about biased with correlated features since all your feature are highly correlated. In your case, all measurements at 19 (Tb, Pr, Fr, and anomalies) are correlated with each other, same for 37. This will affect the feature importance. I would just mention how it would affect it.
It is true, and we have removed the radiometer-derived ratios and anomalies from the RF approach, also as they do not contribute to the results.

Figure 5: Y axis are never the same. It's hard to distinguish differences between each cluster.
It is rather difficult to completely normalise the y-axes, because doing this would make all dry-snow regions appear as a straight line. We have therefore used a same range of y-axis for dry regions, and another range of y-axis for melt regions. Due to the suggestion of Referee 2, we have moved this figure to Appendix B.

Line 358-359: Again correlation between derived sat (PR, FR) and the Tb at 19,37 is probably high. So the importance of Pr and Fr is reduced. Most of the information contain in Pr is probably also in the Tb.
They have been removed from the RF approach.

Figure 7: the caption is missing a parenthesis at the end.
This has been corrected in the revised manuscript.

Line 393: Missing a come between high and close? "the RMSE between IMAU-FDM and RF is high close to Vinson Massi".
This has been corrected in the revised manuscript.

Figure 10: Can you change the bright yellow color? The label is unreadable.
Since the comparison of density variations with polarisation ratios is not valid anymore, this figure has been removed in the revised manuscript.

Line 413: I think that is already well established when you look at snow density, microwave data and radiative transfer... (Brucker et al 2010, 2011 . Picard et al 2009, 2014). I would reword…
Agreed. "Our findings reveal" has been changed to "out study is based on" in the revised manuscript.

Line 454: Grain size also influence Tb. Perhaps some words on how it affect the final result?
It is true and is the main reason why we assumed that a non-linear machine learning model could account for the effect from grain size and other drivers. However, due to the lack of information on grain size, we cannot arrive at a better (or more solid) conclusion regarding how it affects the final result.